DOCUMENT RESUME

ED 321 729 IR 014 104

AUTHOR Schmitt, Dorren Rafael
TITLE Methodological Weaknesses with CAI Research.
PUB DATE 24 Feb 89
NOTE 19p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Savannah, GA, February 22-25, 1989).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Instruction; *Courseware; Evaluation Criteria; Instructional Effectiveness; Meta Analysis; *Research Design; Research Methodology; *Research Problems; Sample Size; Statistical Analysis

ABSTRACT

Since about 1962, researchers have explored the effectiveness of educational software in the schools under various conditions, and many have concluded that computer assisted instruction (CAI) is effective. Many of the studies reviewed to support this conclusion, however, had design flaws. Reviews of educational software effectiveness studies have noted four major research design weaknesses: (1) small sample sizes; (2) lack of identified criteria for determining "quality" software; (3) inappropriately used statistics; and (4) inadequate time allocated to conducting the study. Because the actual studies have design flaws, these research design weaknesses will be reflected in the results of the studies and even compounded should meta-analyses using these studies be performed. These major design weaknesses are discussed and selected studies are reviewed to illustrate them. Practical solutions to the design flaws are proposed, and the effects of the design flaws on meta-analyses are outlined. An appended table displays the 16 randomly-selected studies used to illustrate the design flaws together with data for each one on sample size, duration of the study, the software used, and the statistical techniques used to analyze the data. (45 references) (GL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED321729

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

METHODOLOGICAL WEAKNESSES WITH CAI RESEARCH

Dorren Rafael Schmitt

University of New Orleans

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Dorren Rafael Schmitt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

IR014104

Paper presented at the annual meeting of the Eastern Educational Research Association, Savanah, GA, February 24, 1989.

METHODOLOGICAL WEAKNESSES WITH CAI RESEARCH

The 1960s was the initial introduction of computers into education. Pioneered by IBM and Philco-Ford, computer-assisted instruction (CAI) was developed and brought computers closer to being in the classroom. In the late 1970s, inexpensive computers, such as the Apple II, TRS-80 model I, and Commodore PET, were available for K-12 schools. This began a trend toward developing vast numbers of CAI programs and CAI research.

Since about 1962, researchers have explored the effectiveness of educational software in the schools under various conditions. Zinn (1965), Wold and Hunter (1984), and James (1986) have argued that CAI is effective for instructing students. But, many studies that these authors reviewed to draw their conclusions had design flaws. In addition, meta-analyses have been performed which contend that CAI is at least not harmful and probably effective (Burns & Bozman, 1981; Hartley, 1978; Hasselbring, 1986; Kulik, Bangert, & Williams, 1983).

Reviews of educational software effectiveness (Chambers & Sprecher, 1980; Colorado, 1988; Schmitt, 1987) studies have noted four major research design weaknesses: small sample sizes, lack of identified criteria for determining "quality" software, inappropriately used statistics, and inadequate time allocated to conducting the study. Because the actual studies have design flaws, these research design weaknesses will be reflected in the results of the studies and even compounded should meta-analyses, using these studies, be performed since meta-analyses of an analysis of data.

This paper discusses the four major research design weaknesses, reviews research studies to illustrate these weaknesses, and discusses some practical

solutions to the design flaws. Affects on meta-analysis will also be discussed.

Selection of Research Studies

Several sources of educational papers were surveyed, ERIC, Dissertation Abstracts Internation, and educational journal, from 1970 through 1987. Sixteen CAI studies in mathematics published during this 17 year period were randomly selected to be used to illustrate the research design weaknesses in CAI studies.

The appendix has a table lists the various characteristics of CAI studies in mathematics. This table present the sample size, the length of the study, the software used (commercial, public-domain, self-made, or unknown), the statistical procedures employed, and the conclusions (more effective, no difference, less effective) with traditional instruction.

Small Sample Sizes

One controllable research design aspect is the sample size. The second level questions the researcher asks are:

Who is my population?

Who will be in my sample?

And, how many people will be in my sample?

These three questions usually follow the definition of the research problem and development of the research questions. Tables and/or formulas can be found in any number of social science statistics books to determine the sample size required to have sufficient "power" for the study.

CAI researchers like many other educational researchers neglect sample size. The mean sample size in these 16 CAI studies reviewed was 74. This is the sample size for the entire study, not just a level of a way. Difficulty achieving statistical significance is surprising due to this fact. Most CAI studies do not have sufficient "power". Not only are these studies not statistically different, but if educational significance is explored, through effect sizes, many of these same studies do not indicate educational significance in the findings. This is also supported by conclusions by (Colorado, 1988; Schmitt, 1987).

As examples of small sizes, Pachter (1979) used a sample of 29. Other studies suing a sample size in the thirties were Sterrett and Karian (1978), Lang (1973), Robichaud (1986), Imboden (1986), and Wilson (1986) with samples between 29 and 38. Three studies did have large samples: Goodson (1975) with 93, Sherman (1975) with 117, and Shapiro (1977) with 400 subjects.

There are solutions to this problem. Instead of using a single school and one grade level, the researcher could use two schools. One school could be the control group and the other could be the experimental group. If more than two groups are needed expand the number of schools. This would increase the sample two fold or by the number of researcher groups needed for the study.

A second solution would to take multiple schools and have all of the research groups microcosms of the research study. Instead of having three schools, each school representing a different group, have each school have the three groups thus amounting to a similar sample size as the above suggestion.

A final suggestion would have several researchers from different school districts or states work on the same study. Each researcher would have all the

groups in his "sub-study". This would not only increase the sample size but also the generalizability of the study.

These solutions are just three logically throughout solutions to the problem of sample size. This author feels that one of the major reasons statistical and/or educational significance has not been found is due to sample size. The lack of greater effectiveness over traditional instruction has hindered the expanded use of CAI in some form of classroom instruction. The findings of these past studies has also hindered the further investigation of CAI effectiveness similar to the past and in other aspects.

Duration of the Study

The duration of the study is usually largely under the control of the researcher. Although, the availability of students and the liberties schools and teachers give researchers may be constraining, the researcher does have a reasonable amount of control over how long the study will take place.

The second design flaw found in many CAI research studies is the length of the study. Several of the studies (i.e., Imboden, 1986; Shine, 1983; Pachter, 1979; Goodson, 1975) only lasted through one topic. In science and mathematics, a topic may last one to three days. This limits the study to the length of time the single topic is discussed. Other studies lasted through multiple topics. Finally, there were some studies that lasted a semester or longer: Sterrett and Karian (1978), Lang (1973), Crenshaw (1983), and Ozarowski (1973).

The length of time is largely determined by the number of software packages are used in the study. On one dimension there are the topics that will

be included in the study (i.e., solving linear equations, long vowels, the solar system). On another dimension, usually neglected, is the variety software package types that can be used (i.e., drill and practice, tutorial, simulation and demonstration, games). The studies report the topics and usually only employ one type of software package. A more accurate deception of how effective CAI is on a given topic might be to use several different software types on the same students or on different student and observe differences.

Even those studies which lasted a semester, the students may have only had limited use of the computers. Use once a week or every other week could possibly not show differences as much as using the computers twice or three times a week throughout the semester.

This is not to say that the general duration, a semester or even multiple topics is not long enough, but this author is suggesting that actual computer usage would be a better indicator. CAI studies have varied from students using the computer 15 minutes per session to one hour per session. The duration of the sessions makes a dramatic difference when discussing ten 15 minute sessions compared to ten one hour sessions.

Additionally, if the sessions are distributed over various topics and multiple software types are used, there is more likely to be statistical and/or educational significance. The length of the study most definitionally has an effect on either measure of significance. And obviously, the results will also be more stable over time and over software types. This may also produce more generalizable results.

Software Quality

The lack of knowledge of the software quality or the lack of quality software used is considered an inexcusable design flaw. Educational software quality ranges from abonable to excellent (Allen, 1984; Bitter, 1984; Burke, 1982; Jensen, 1985; Kamil, 1984; Komoski, 1987; Lathrop & Goodson, 1983; Staples, 1985; Taylor & Ely, 1987). Staples (1985) has stated that there is a tremendous amount of variation in the quality of software, but the largest quantity is the poorer software. With the amount of poor quality software, the "educational merits of [software] packages need to be questioned" (Niman, 1987, p.57).

All of the studies reviewed either did not state what software was used or defined the software as made by the researcher (self-made). There was no discussion in any study pertaining to the quality of the software. This is disturbing since one major criticisms of CAI by teachers is that "there is no good software our there" (Balajthy, 1988, p.51).

The quality of the software is the backbone of all CAI studies. If the software is not excellent or at least good, then CAI may be found to be ineffective or at least as effective as "traditional" instruction. The quality or Jack there of in the research software is compounded by the previous two flaws, particularly the duration of the study.

The major concern for this author is not so much that the researchers developed the software for the study, but that the quality of the software was not documented. This is particularly a concern in those papers from ERIC since there is no page limitations.

Suggestions for documenting the quality of software whether it may be commercial, public-domain, or self-made. First, it the software being used in the study was commercially produced, check in ERIC or subject matter journals for reviews of the software (i.e., MicroSIFT, Alberta, Arithmetic Teacher, Mathematics Teacher, English Teacher, Reading Teacher, Science Teacher). Second, if the software was not commercially produced, locate a software evaluation instrument (i.e., MicroSIFT, NCTM, NCTE, or other published in ERIC or various journals). Third, provide documentation from referencing the review found or a summary of the software evaluation. Finally, discuss reason for using particular software particularly if the software was self-made and/or received a poor review or evaluation.

It is this authors opinion that the major downfall of past CAI research studies has been the lack of quality software. This has placed both CAI research and the use of CAI software in jeopardy. Without credible software research studies will fail to accurately determine effectiveness or ineffectiveness of CAI.

Statistical Procedures Employed

To this point, most studies have relied on univariate statistics. The predominant statistical procedures employed have ben analysis of variance (ANOVA) and t-tests. In and of themselves there is nothing wrong with these statistical procedures, but they have been misused and overused.

The three statistics that were used by all of the reviewed studies were ANOVA, t-test, and regression. Given the data, the general type of statistics used was appropriate. One problem is that in several studies, multiple ANOVAs and

multiple t-tests were employed. This is problematic.

The same logic used by textbooks supporting the use of ANOVA tests over t-tests can be made for using multivariate statistics over multiple ANOVAs. If the alpha is set at .01, with one ANOVA there is a 1 percent chance that the test would be declared statistically significant (Asher, 1976). This is termed experimentwise Type I error.

When more than one ANOVA, or any other univariate test, is performed on the same dependent variables the experimentwise Type I error probability is increased exponentially. This error can range from the set alpha level, when all of the dependent variables are perfectly correlated, to "1 - (1 - alpha) raised to the K power, where K is the number of statistical test conducted" (Thompson, 1986, p.6), when all of the dependent variables are perfectly uncorrelated. The most troubling fact is that except at the two extremes, the actual experimentwise error rate cannot be readily determined (Thompson, 1986). The result is a greater probability of statistical significance (McMillan * Schumacher, 1984; Sowell & Casey, 1982). Thus, instead of multiple univariate statistics, one multivariate statistic should be employed.

But beyond the fact that the probability of statistical significance is greater, multivariate statistics also better represent the reality of the classroom. And, the classroom is where the CAI will be used. Thompson (1986) argues that the statistical procedures should be as complicated as the environment in with the data was collected. The classroom is not simple. And, there is not one dependent variable that is involved. Achievement, interest, and attitude are just three aspect that are involved in all learning, whether or not the computer is used. These

aspects, and possibly others, need to be assess together and not independently. Additionally, different aspects of achievement should not be investigated independently, but together as well.

CAI research, like much of educational research, has neglected the need and usefulness of multivariate statistics. Univariate statistics can be used. And there are appropriate situations. But, when multiple univariate tests need to be employed, multivariate tests are then most appropriate.

Finally, educational significance or effect sizes has been neglected throughout CAI research. An effect size, often $\underline{R}$, is used to determine educational significance. One of the easiest effect sizes to calculate is (sum of square explained divided by sum of squares total). Even if there is not statistical significance, especially with a small sample size, there may be educational difference. And obviously, the higher the percentage of difference the greater the difference.

Design Flaws Affecting Meta-Analyses

Simply put, meta-analysis is the "analysis of analyses" (Glass, 1976), p.3). Basically, meta-analysis is the reviewing of a collection of studies to summarize statistics (Kulik & Kulik, 1988). Kulik and Kulik (1988) explain that meta-analysis is not based on statistical significance by effect sizes and the strength of relationships.

Because meta-analyses rely on other studies, the meta-analysis study incorporates any problems the reviewed studies might have. Thus, CAI meta-analyses also incorporate the flaws described above. Most specifically, length of the study and the quality of the software. Due to the use of effect sizes, the

sample size does not influence the meta-analysis as much as if statistical significance was investigated, since statistical significance is largely a product of sample size.

The flaws of the studies do influence the results of meta-analyses. This compounds the problems since many educators rely on the meta-analyses as the literature for the effectiveness of CAI because of the vast numbers of CAI studies. The meta-analyses can provide a false sense of success with CAI that has not been shown to necessarily be present in CAI studies.

Summary

There are design flaws with past CAI effectiveness studies. After reading a few studies this conclusion seems obvious. But, all is not lost. If researchers need the recommendations for improving CAI studies there may be some hope.

As the technology of computers grows and the demand for computers in schools from within the education profession and from parents increases, there needs to be accountability that CAI is useful, effective, and necessary. Unless there are sound research studies to backup the use of CAI, it will fall like a dead duck.

CAI may not be effective as the sole means of instruction, but that does not mean it is not useful and necessary. The CAI researchers need to provide the fuel to spark the demand for quality CAI and appropriate uses of the software.

REFERENCES

Allen, M. W. (1984). Educational computing software: State of the art in 1984. Educational Technology, 24, (10), 7-10.

Asher, J. W. (1976). Educational research and evaluation methods. Boston: Little, Brown.

Balajthy, E. (1988). Computers and instruction: Implications of the rising tide of criticism for reading education. Reading Research and Instruction, 28, 49-59.

Barnes, K. (1984). Selecting 'good' stuff in computer training in a 2-sided job. PC Week, 1(10), 23.

Bitter, G. G. (1984). Hardware and software selection and evaluation. Computers in the Schools, 1(1), 13-28.

Burke, R. L. (1982). CAI sourcebook. Englewood Cliffs, N.J.: Prentice-Hall.

Burns, P. K., & Bozeman, W. C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship? Educational Technology, 21(10), 32-39.

Chambers, J. A., & Sprecher, J. W. (1980). Computer-assisted-instruction: Current trends and critical issues. Communications of the ACM, 23, 332-342.

Colorado, R. J. (1988). Computer-assisted instruction research: A critical assessment. Journal of Research on Computing in Education, 20, 226-233.

Crenshaw, H. M. (1983). A comparison of two computer assisted instruction management systems in remedial mathematics. Dissertation Abstracts International, 43, 3467A (University Microfilms No. 83-06-839).

Enochs, J. R., Handley, H. M., & Wollenberg, J. P. (1986). Relating learning style, reading vocabulary, reading comprehension, and aptitude for learning

achievement in the self-paced and computer-assisted instruction modes. The Journal of Experimental Education, 54, 135-139.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5, 3-8.

Goodson, C. E. (1975). A study of the effectiveness of computer assisted instruction as compared to traditional instruction when utilized in technical mathematics for college students in business and engineering technology. Dissertation Abstracts International, 36, 2688A-2689A. (University Microfilm No. 75-23,934).

Hartley, S. S. (1978). Meta-analysis of the effects of individually paced instruction in mathematics (Doctoral Dissertation, University of Colorado, 1977). Dissertation Abstracts International, 38, 4003A. (University Microfilms No. 77-29,926).

Hasselbring, T. S. (1986). Research on the effectiveness of computer-based instruction: A review. International Review of Education, 32, 312-324.

Imboden, H. H. (1986). The effects of computer enhanced instruction in teaching concepts of percent to low achieving college students. Dissertation Abstracts International, 46, 3378A. (University Microfilm No. 85-25,768).

James, E. B. (1986). Computer-based teaching for undergraduates: Old problems and new possibilities. Computer Education, 10, 281-292.

Jensen, C. B. (1985). Using what we have learned in the past to improve the future of courseware design. AEDS Journal, 19,(1), 28-48.

Kamil, M. L. (1984). Software: Evaluation and decisions. Computers, Reading, and Language Arts, 1(4), 15-16, 26.

Komoski, P. K. (1987). Keynote address: Educational microcomputer software evaluation: Its challenge and its uniqueness. In J. Moonen and T. Plomp (Eds.), Eurit 86 developments in educational software and courseware (pp. 399-404). Oxford: Pergamon Press.

Kulik, J. A., Bangert, R. L., & Williams, G. W. (1983). Effects of computer-based teaching on secondary school students. Journal of Educational Psychology, 75, 19-26.

Kulik, J. A., Kulik, C-L. C. (1988). Meta-analysis: Historical origins and contemporary practices. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Lang, M. T. (1973). Computer extended instruction in introducing calculus. Dissertation Abstracts International, 34, 5662A. (University Microfilm No. 74-5, 271).

Lathrop, A., & Goodson, B. (1983). Courseware in the classroom selecting, organizing, and using educational software. Menio Park, CA: Addison-Wesley.

McMillian, J. H., & Schumacher, S. (1984). Research in education: A conceptual approach. Boston: Little, Brown.

Niman, J. (1987). Precollege mathematics software: Limitations and Potential. In J. Moonen and T. Plomp (Eds.), Eurit 86 developments in educational software and courseware (pp. 55-65). Oxford: Pergamon Press.

Ozarowski, P. C. (1973). A study in the design and implementation of a course in the basic fundamentals of statistics via a computer. Dissertation Abstracts International, 34, 2310A. (University Microfilm No. 73-27,317).

Pachter, S. (1979). A computer assisted tutorial module for teaching the factoring of second degree polynomials to regent level ninth year mathematics students. Dissertation Abstracts International, 40, 1843A. (University Microfilm No. 79-23,610).

Robichaud, K. K. (1986). The use of computer assisted instruction with adult basic education students: A comparative study. Dissertation Abstracts International, 47, 1985A. (University Microfilm No. 86-21,681).

Rockhill, T. D. (1971). The development of an individualized instructional program in beginning college mathematics utilizing computer based resource units. Final report. (ERIC Document Reproduction Service No. ED 053 966).

Schmitt, D. R. (1987). The effects of various approaches to using CAI to teach mathematics: A literature review and possible research agenda. Paper presented at the annual meeting of the Mid-South Educational Research Association, Mobile, AL.

Schoen, H. L. (1972). A comparison of four types of feedback to student responses on a CAI unit designed to teach the concept of function. Paper presented at the National Council of Teachers of Mathematics Annual Meeting, Chicago, IL. (ERIC Document Reproduction Service No. ED 061 081).

Shapiro, L. (1977). Tutor resources in mathematics. American Mathematical Monthly, 84, 476-481.

Sherman, J. F. (1975). The development of a computer assisted math review for physical science survey students at Bevard Community College. Ed.D Dissertatio, Nova University. (ERIC Document Reproduction Service No. ED 137 084).

Shine, S. S. (1983). A comparison of programmed instruction versus lecture-demonstration as a method for teaching digital computer arithmetic at the post-secondary school level. Dissertation Abstracts International, 43, 3529A. (University Microfilm No. 83-06, 819).

Sowell, E. J., & Casey, R. J. (1982). Research methods in education. Belmont, CA: Wadsworth.

Staples, B. (1985). Educational computing: Where are we now? Creative Computing, 11(4), 62.

Sterrett, A., & Karian, Z. A. (1978). A laboratory for an elementary statistics course. American Mathematical Monthly, 85, 113-116.

Taylor, R., & Ely, D. P. (1987). Courseware quality: Do development and design procedures make a difference? In J. Moonen and T. Plomp (Eds.), Eurit 86 developments in educational software and courseware (pp. 405-410). Oxford: Pergamon Press.

Thompson, B. (1986). Two reasons why multivariate methods are usually vital: An understandable reminder with concrete examples. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN.

Walker, E., & Azumi, J. E. (1985). The impact of computer assisted instruction on mathematics learning gains of elementary and secondary students. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 275-487).

Wilson, S. A. (1986). The construction of computer-assisted instruction: An application to remedial college algebra. Dissertation Abstracts International, 46, 2222A. (University Microfilm No. 85-23,769).

Wold, A. L., & Hunter, C. B. (1984). The not quite twenty-year history of educational software. Media and Methods, 37, 24-25.

Zinn, K. L. (1965). Computer assistance for instruction. Automated Education Letter, 1, 4-13.

APPENDIX

TABLE OF STUDIES

Table 1. Mathematics Studies

Author(s)	Sample Size	Duration of Study	CAI Software	Statistics
Rockhill (1971)	53	4 topics	self	T-tests
Schoen (1972)	60	Semester	self	ANOVA
Lang (1973)	34	Semester	self	ANOVA
Ozarowski (1973)	46	Semester	self	T-test, ANOVA
Goodson (1975)	93	1 topic	self	ANOVA
Sherman (1975)	117	Semester	self	ANOVAs
Shapiro (1977)	400	Semester	self	ANOVA
Sterrett & Karian (1978)	35	Semester	self	ANOVA
Pachter (1979)	29	1 topic	self	T-test
Crenshaw (1983)	43	Semester	CCC	T-tests
Shine (1983)	41	1 topic	self	ANCOVA
Walker & Azumi (1985)	153	5 months	MATH1-6 program	T-tests, regres.
Enochs & others (1986)	50	Semester	self	ANOVA, Pearson
Imboden (1986)	38	1 topic	self	ANOVA
Robichaud (1986)	32	Semester	self	T-tests
Wilson (1986)	38	6 topics	self	T-tests

mean sample size 79; standard deviation 92.3
mean sample size without Shapiro 57; standard deviation 35.8

Note. CCC is Computer Curricular Corporation System